DOCUMENT RESUME

ED 346 130 TM 018 420

AUTHOR Abdel-fattah, Abdel-fattah A.
TITLE Comparison of the Chi-Square Procedure with the Symmetric Regression Procedure for Developing a Common Metric in Item Response Theory.
PUB DATE 92
NOTE 12p.; Paper presented at the Joint Annual Meeting of the Psychometric Society and the Classification Society (New Brunswick, NJ, 1991).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Chi Square; Comparative Analysis; *Equated Scores; *Estimation (Mathematics); *Item Response Theory; Least Squares Statistics; Literature Reviews; Maximum Likelihood Statistics; *Regression (Statistics); *Scaling; Weighted Scores
IDENTIFIERS Linear Equating Method; Outliers; *Symmetric Regression Procedure

ABSTRACT

A modified regression procedure for finding a common metric for item response theory (IRT) parameter estimates and for total test scores is proposed and compared to the chi-square method of equating. The chi-square method was chosen after literature review for the currently used scaling procedures. The modification used with the regression analysis was to account for its asymmetry. The new method was compared to the chi-square procedure and was found to produce similar transformation constants. The new method is simpler and more flexible than the chi-square procedure, and it is inexpensive and conceptually valid. It has been implemented in a computer program that equates the scores of two tests. Parameter estimates from different calibrations or different tests can also be scaled to a common metric by using this program. It can be used for linear, non-linear, raw score, or true score equating. The equating can be performed by using the ordinary least squares, the weighted least squares, or the maximum likelihood estimation procedure. The weighting is used to account for outliers. Weighting can also be performed by using the reciprocal of the average standard deviations of the two variables to be equated. The program used in this study is available in both the Statistical Analysis System (SAS) and FORTRAN. One table and 12 graphs present study data, and there is a 10-item list of references. (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED346130

Comparison of the Chi-square Procedure with the Symmetric Regression Procedure for Developing a Common Metric in Item Response Theory[1,2]

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

ABDEL-FATTAH A. ABDEL-FATTAH

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Abdel-fattah A. Abdel-fattah

Educational Testing Service, 1992

[1] The opinions expressed in this paper are those of the author and should not be mistaken to represent the official policy of the Educational Testing Service.

[2] This paper was presented for the 1991 joint meeting of Psychometric Society and Classification Society held at Cook Campus of Rutgers University.

TM018420

BEST COPY AVAILABLE

Abstract

A modified regression procedure for finding a common metric for item response theory (IRT) parameter estimates and for total test scores is proposed and compared to the chi-square method of equating. The chi-square method was chosen after literature review for the currently used scaling procedures. The modification used with the regression analysis was to account for its asymmetry. The new method was compared to the chi-square procedure and was found to produce similar transformation constants. The new method is simpler and more flexible than the chi-square procedure. It is also inexpensive and conceptually valid. It has been implemented in a computer program which equate the scores of two tests. Parameter estimates from different calibrations or different tests can also be scaled to a common metric by using this program. It can be used for linear, nonlinear, raw score, or true score equating. The equating can be performed by using the ordinary least squares, the weighted least squares, or the maximum likelihood estimation procedure. The weighting is used to account for outliers. Weighting can also be performed by using the reciprocal of the average standard deviations of the two variables to be equated. The program used in this study is available in both SAS and FORTRAN.

Review

To obtain metrics common to two distributions, Marco (1977) found the slope A and the intercept B so that the mean and standard deviation of the transformed distribution equal to the mean and the standard deviation of the estimated distribution (e.g., item difficulties). In this method, sample moments, can be seriously affected by poorly estimated item difficulties. Stocking and Lord (1983, p. 203) stated that Cook, Eignor, and Hutton (1979) have attempted to solve this problem by restricting the range of item difficulties. Linn, Levine, Hastings and Wardrop (1980) employed weighted moments to reduce the influence of these outliers. All points with the same standard error are treated in the same way regardless of status of their outliers.

Stocking and Lord (1983, p. 203) reported that Bejar and Wingersky (1981) used a robust method that gives smaller weights to the outlying points. They treated all outliers in the same fashion regardless of their standard errors. Stocking and Lord (1983) presented a method to overcome potential problems of the procedures by Linn et al. and those by Bejar and Wingersky. This new method gives low weights to poorly estimated parameters and to the outliers. A drawback of this method is that only the information contained in b parameters is used whereas the information in a parameters is ignored. Stocking and Lord calculated A and B by minimizing the mean square difference between two test characteristic curves (TCC). This method includes the information in a and b parameters in the calculation but it does not take into account the standard errors of estimates. The chi-square method, by Divgi (1985), is simpler and cheaper than the TCC method. It makes use of the information about sampling error, a feature not accounted for in the TCC method. The chi-square method was used in this paper because of this feature.

Stocking and Lord (1983) stated that the problem of finding the linear relationship between two sets of numbers might seem to call for simple regression techniques. They also indicated that simple regression is an incorrect approach because it assumes that the independent variable (estimated item difficulties or abilities from one calibration) is measured without error, which is not the case for parameter estimates, and that regression is not symmetric with respect to treatment of dependent and independent variables.

von Eye and Rovine (1992) provided a comprehensive review for symmetrical regression procedures; this review is summarized below. The earliest work on symmetrical regression was presented by Pearson (1901) and the only application at that time was by McDowell (1901) who also contributed the first application to the principal component solution for regression asymmetry. Recent alternative to the solution by Pearson are a) double regression which is the bisector of Y-X and X-Y regression lines (Rubin, Burstein, & Thonnard 1980), b) impartial regression which is the geometrical mean of slopes of Y-X and X-Y regression lines (Stromberg, 1940), c) robust ordinary least square (OLS) after excluding outliers (von Eye & Rovine, 1991) which was based on least median of squares by Rousseuw (1984), and d) measurement error regression model or the non-OLS solution (Isobe, et al., 1991). The differences between the first two alternative are never large except when one of the lines is inestimable. The third solution is most appropriate if one is concerned about the adverse effect of the outliers on the

analysis. The fourth alternative incorporate knowledge concerning the structure of measurement error which stems from, and consequently more suitable for, experimental results.

The problems of asymmetry and measuring the independent variable without error can thus be solved by performing one of the solutions mentioned above. The solution adopted in this paper is the arithmetic average slope of the Y-X and X-Y slopes and that of Y-X and X-Y intercepts as the new slope and intercept of the transformation procedure. This approach is most similar to the bisector line solution by Rubin, Burstein, and Thonard (1980). Thus, the simple regression analysis using the first calibration as the dependent variable and the second as the independent variable, then by using the first calibration as the independent variable and the second as the dependent variable. The average of the two slopes and the average of the two intercepts are the transformation constants. The averaging process will cause the transformation line to bisect the scatterplots of the equated estimates from the two calibrations. This way the transformation constants are not biased for one calibration over another. Divgi (1985) indicated that there is no theoretical criteria for choosing among different methods, or for evaluating the quality of a particular method. The transformation line should bisect the scatterplot of the two calibrations (Stocking and Lord, 1983).

This symmetric regression is an appropriate procedure when the variables represented on both axes are subject to measurement error. That is, neither can be considered the dependent or the independent variable. This is actually the case in test equating in which the magnitudes of errors associated with the two tests are similar and the conversion line from the first test to the second is the same as that for the second to the first. By using this procedure we can equate several tests simultaneously, and we can test for curvilinear relationship(s) and for outliers. If the relationship between the two calibrations is nonlinear we can fit a nonlinear regression and estimate the transformation constants more accurately than if we used linear regression. After detecting the outliers, they can be excluded, substituted for the average observation or a randomly selected observation. More accurate transformation constants are obtained by weighting the calibrations with the reciprocal of their variance to give a less weight to the inaccurately estimated parameters or outliers. Any of these treatments to the outliers will lead to more stable transformation constants than if we do not treat them.

Data and Methods

The binary scores were generated for 20 items and 250 examinees. The item and the ability parameter estimates were calculated from the binary scores using BILOG's marginal maximum likelihood (MML) and LOGIST's joint maximum likelihood (JML) procedure. The estimates from the two calibrations were equated by the regression approach and by the chi-square approach. The chi-square procedure minimizes a quadratic function which equals the chi-square statistic for item bias (Lord, 1980). It includes information about estimation errors of the discrimination and difficulty parameters.

The regression procedure used the b estimates to calculate the transformation constants. In spite of the wide range of variables that can be handled by this regression procedure, only one set of data were used. The variables of this data set are linearly related and appeared to be free from serious outliers. Therefore, no outliers analysis was necessary and the linear regression was used.

The transformation constants A and B were calculated for the two scaling procedures. These are the symmetric regression procedure and the chi-square procedure. The new-to-old regression equating (by using the new form b as the independent variable and the old form b as the dependent variable) was also included in the comparison. The regression procedure was implemented by using the SAS program. The listing of the program can be provided by the author upon request. The chi-square procedure is written in FORTRAN by Divgi (1985). Overall, the symmetric regression procedure provided the

Results, Discussion, and Conclusion

Table 1 shows the slopes and the intercepts produced by the compared procedures. The mean square deviations (MSD) for each transformed parameter estimate is also shown in Table 1. The two procedures appeared to be generally similar. For the a parameter estimates, the chi-square procedure provided the smallest MSD. For the b estimates, the new-to-old regression equating provided the smallest MSD. For the θ parameter estimates the symmetric regression procedure produced the smallest MSD. The two procedures produced similar and satisfactory results. Thus in practical testing situations the use of any of the two procedures is desirable. The 45 degree line bisect the transformed a-, b-, or θ estimates in a similar manner for the three procedures as shown in Figures 1 to 3 for the a-estimates, 4 to 6 for the b-estimates, and 7 to 9 for the θ estimates. For the a-estimates, the symmetric regression transformation was more similar to the new-to-old transformation (Figure 1) and the untransformed a-estimates (Figures 10 and 11); the chi-square transformation was narrower in range than the untransformed and the two transformed a-estimates (Figure 2). For b- and θ-estimates, Chi-square transformation tended to expand the bottom of the scale when compared to the untransformed and the two other transformed estimates. The differences among these procedures are slight for the data set used in this study. To generalize these results various types of data sets (such as non-normal abilities or difficulties and nonlinear relationships) need to be investigated using these procedures.

Insert Table 1 about here

Insert Figures 1 to 9 about here

Table 1. The Slopes and the Intercepts from Symmetric Regression and Chi-square procedure.

Comparison Criteria	New-to-Old Regression	Symmetric Regression	Chi-square
Slope	0.95004802	0.938755	1.180319
Intercept	-.15548261	-0.00240608	-.334128
MSD for a	0.314903	0.0997592	0.0631042
MSD for b	$1.0470 \times e^{-16}$	0.44111	0.697335
MSD for ability	16.3509	2.41552	3.81861

BIBLIOGRAPHY

Divgi, D.R. (1985). A minimum chi-square method for developing a common metric in Item Response Theory. Applied Psychological Measurement, 9, 413-445.

Isobe, T., Feigelson, E. D., Akritas, M. G., Babu, G. J. (1990). Linear regression in astronomy I. The Astrophysical Journal, 354, 104-113.

Linn, R.L., Levine, M.V., Hastings, C.N., & Wardrop, J.L. (1982) Item bias in a test of reading comprehension. Applied Psychological Measurement, 5, 159-173.

Pearson, K. (1901). On the systematic fitting of curves to observations and measurements. Biometrika, 1, 265-303.

Marco, G.L. (1977). ICC solutions to three intractable testing problems. Journal of Educational Measurement, 14, 139-160.

Rousseeuw, P. J. (1984). Least median of squares regression. Journal of the American Statistical Association, 79, 871-880.

Rubin, V. C., Burstein, D., & Thonnard, N. (1980). A new relation for estimating the intrinsic luminosities of spiral galaxies. The Astorphysical Journal, 242, 149-152.

Stocking, M.L., & Lord, F.M. (1983). Developing a common metric in item response theory. Applied Psychological Measurement, 7, 201-210.

von Eye, A., & Rovine, M. J. (1991). Robust symmetrical regression in astronomy. Paper presented at the Conference on Statistical Challenges, in Modern Astronomy, University Park, PA.

von Eye, A., & Rovine, M. J. (1991). Symmetrical regression. Unpublished manuscript.

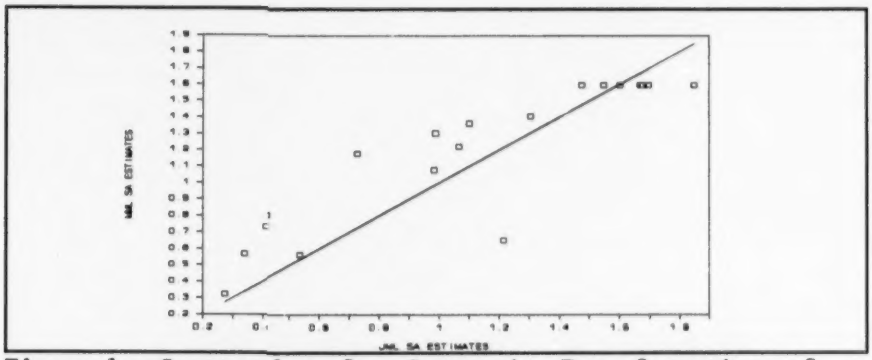

Figure 1. Scatterplot for Symmetric Transformation of a-estimates of MML (SA MML) vs. that for JML (SA JML).

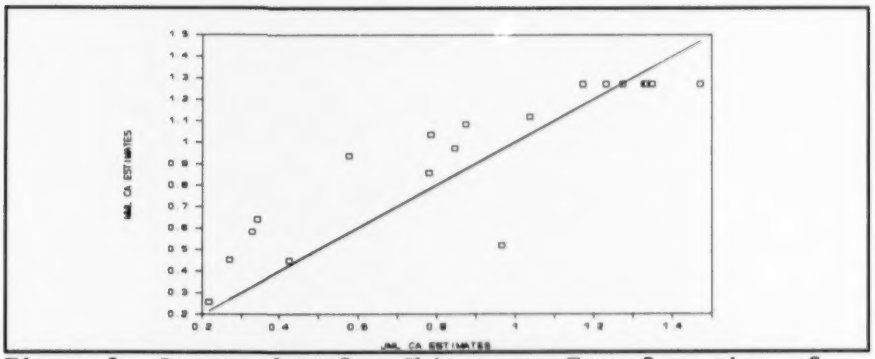

Figure 2. Scatterplot for Chi-square Transformation of a-estimates of MML (CA MML) vs. that for JML (CA JML).

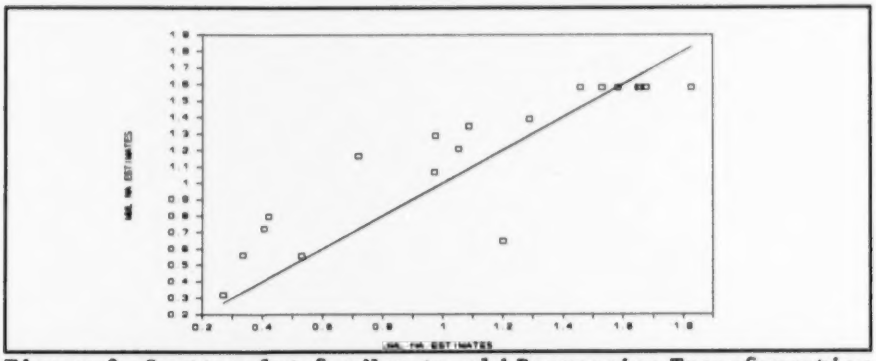

Figure 3. Scatterplot for New-to-old Regression Transformation of a-estimates of MML (NA MML) vs. that for JML (NA JML).

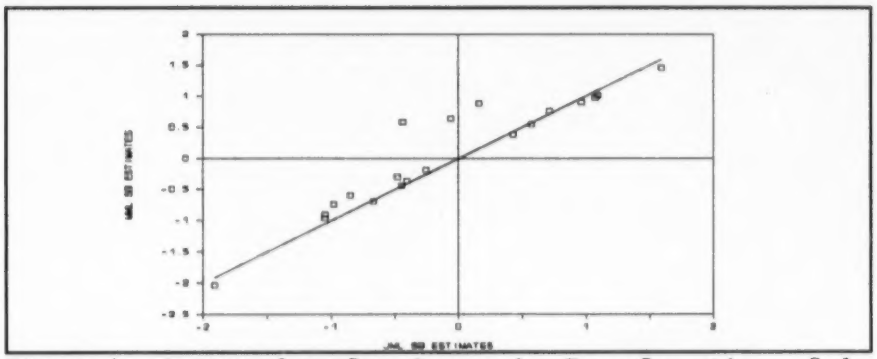

Figure 4. Scatterplot for Symmetric Transformation of b-estimates of MML (SB MML) vs. that for JML (SB JML).

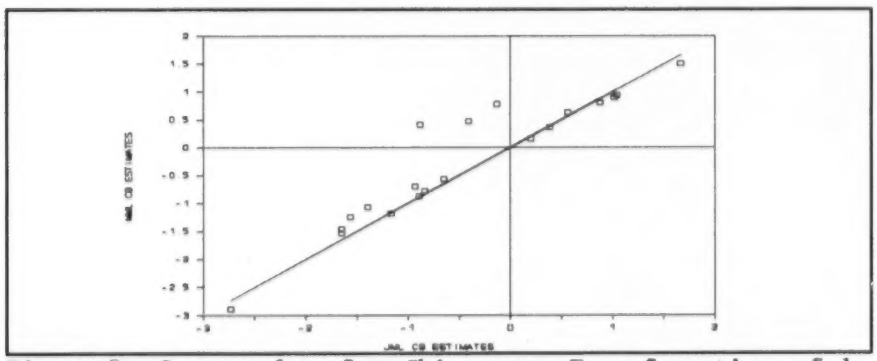

Figure 5. Scatterplot for Chi-square Transformation of b-estimates of MML (CB MML) vs. that for JML (CB JML).

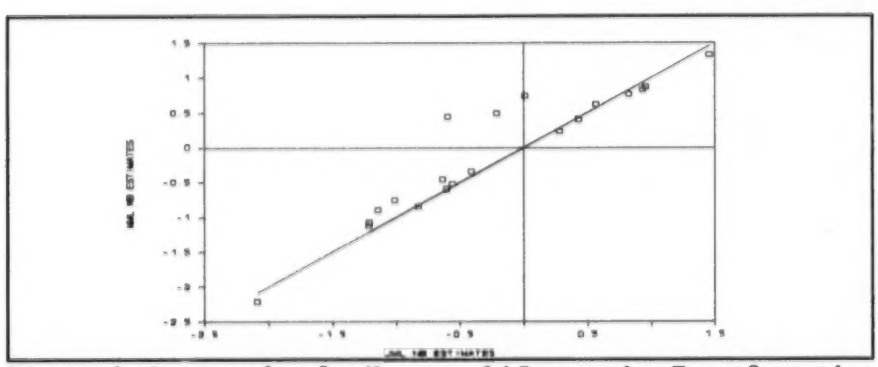

Figure 6. Scatterplot for New-to-old Regression Transformation of b-estimates of MML (NB MML) vs. that for JML (NB JML).

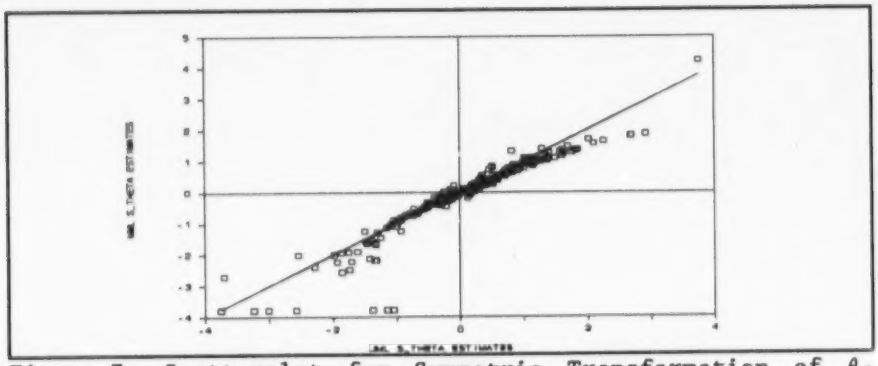

Figure 7. Scatterplot for Symmetric Transformation of θ-estimates of MML (ST MML) vs. that for JML (ST JML).

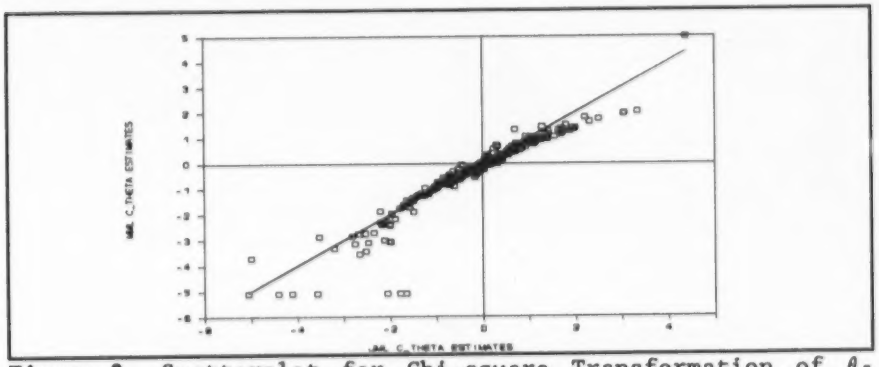

Figure 8. Scatterplot for Chi-square Transformation of θ-estimates of MML (CT MML) vs. that for JML (CT JML).

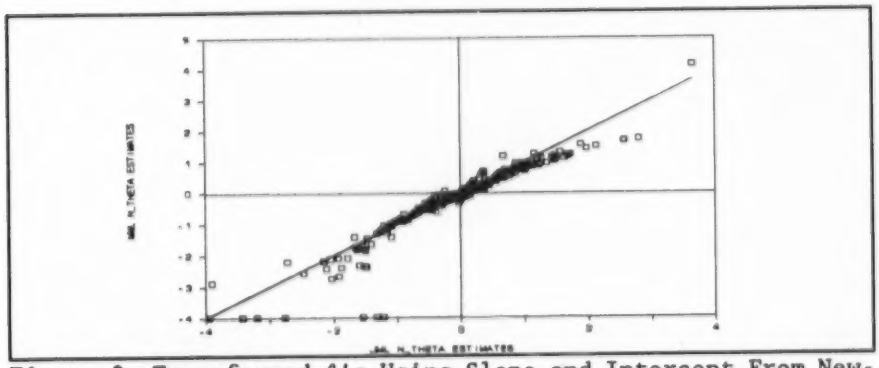

Figure 9. Transformed θ's Using Slope and Intercept From New-to-old Regression
Procedure.

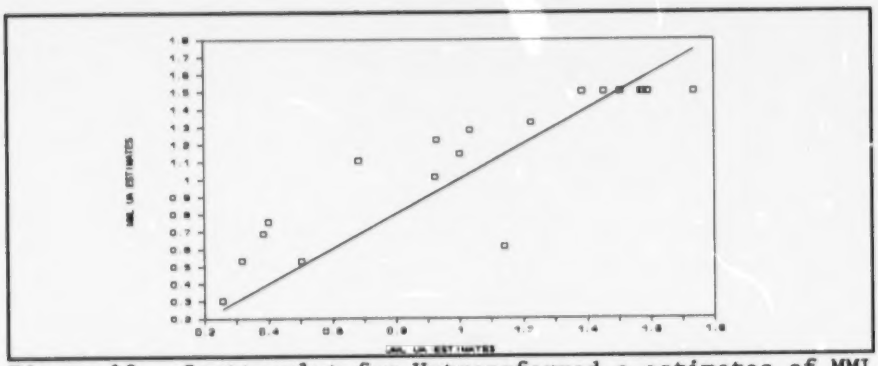

Figure 10. Scatterplot for Untransformed a-estimates of MML (UA MML) vs. that for JML (UA JML).

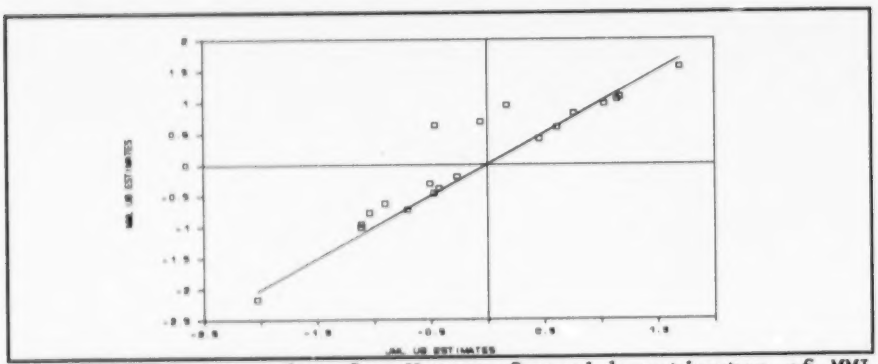

Figure 11. Scatterplot for Untransformed b-estimates of MML (UB MML) vs. that for JML (UB JML).

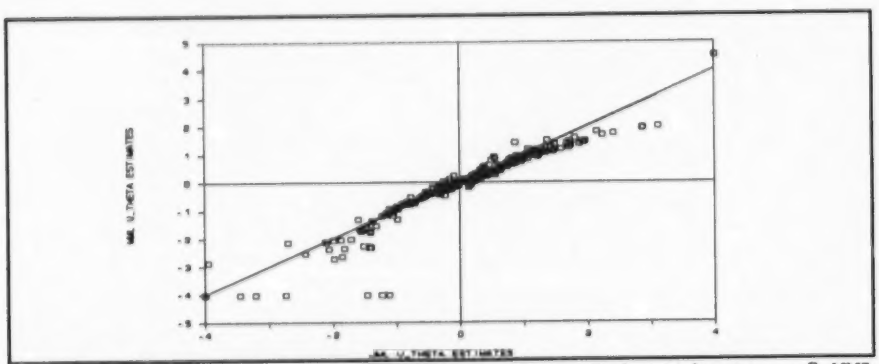

Figure 12. Scatterplot for Untrnasformed θ-estimates of MML (UT MML) vs. that for JML (UT JML).